Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction

Mohna Chakraborty*
Iowa State University
Ames, Iowa, USA

Adithya Kulkarni*
Iowa State University
Ames, Iowa, USA

Qi Li
Iowa State University
Ames, Iowa, USA

KDD-2022

code:
https://github.com/kulkarniadithya/ODAO

Leibniz-Institut für Sozialwissenschaften

Reported by Dongdong Hu

Introduction

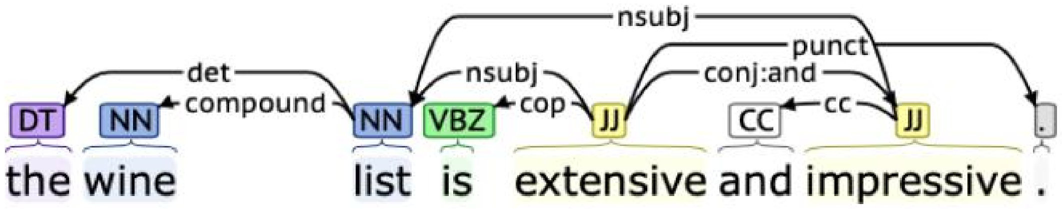


Figure 2: An example of Dependency parse tree

The supervised extraction methods achieve state-of-the-art performance but require large-scale human-annotated training data. Thus, they are restricted for open-domain tasks due to the lack of training data.

$$\mathcal{D}$$

$$R = \{w_1, w_2, ..., w_k\}$$

$$A = \{A_1, A_2, .., A_i\}$$

$$O = \{O_1, O_2, ..., O_j\}$$

$$P = \{(A_i, O_j), ..\}$$

Method

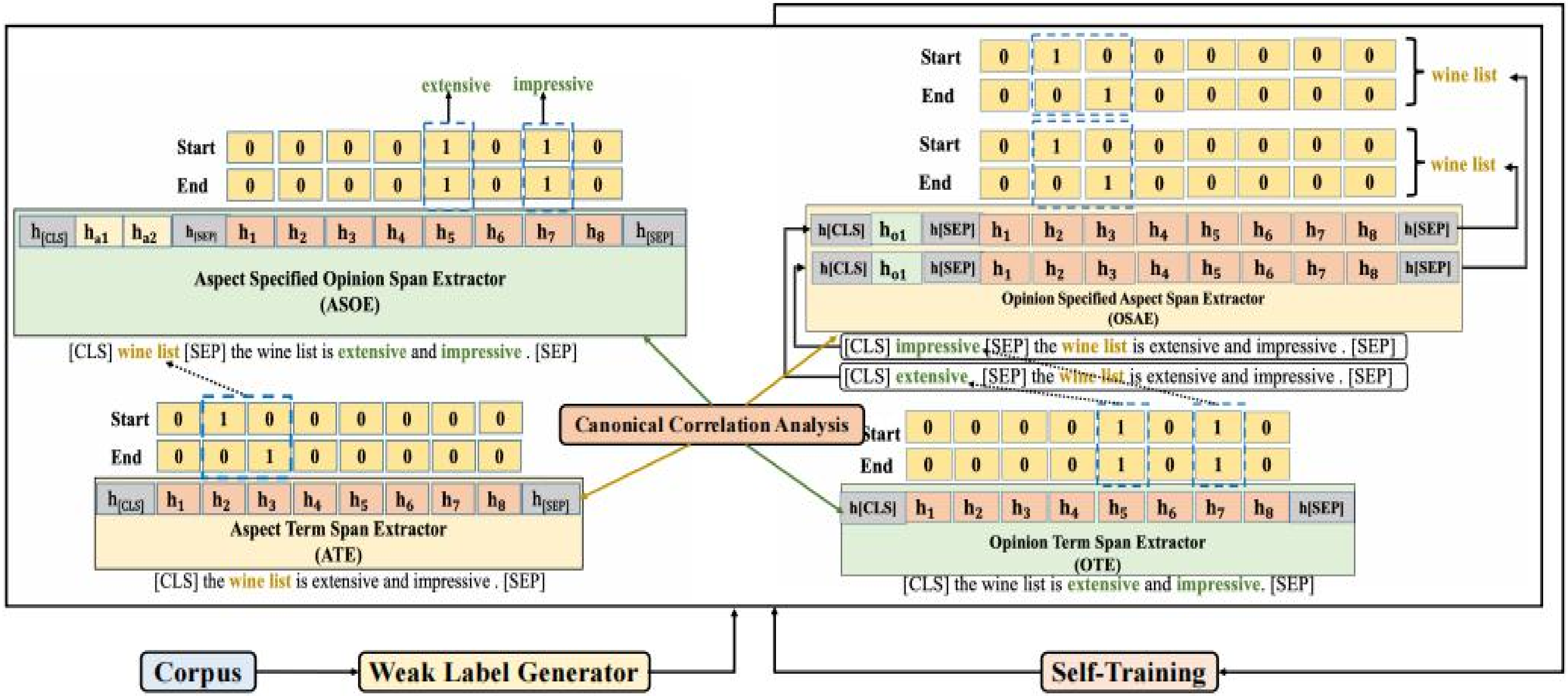


Figure 1: ODAO architecture

Method

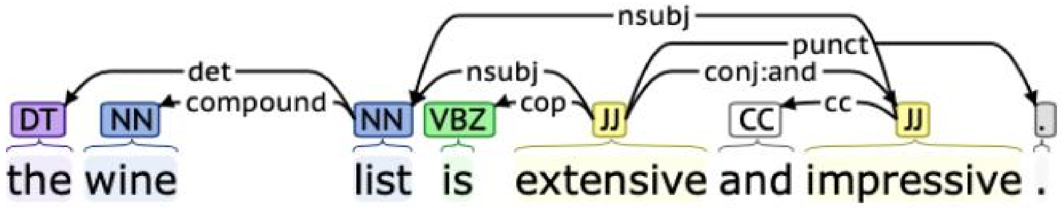


Figure 2: An example of Dependency parse tree

Weak Label Generation

$$AspectTerm = NN \leftarrow nsubj \leftarrow JJ(root) = OpinionTerm.$$

(1) $AT = NN^* \leftarrow nsubj \leftarrow JJ^*(root) = OP$

(2) $OP = JJ^* \leftarrow comp \leftarrow OP$

(3) $AT = NN^* \leftarrow conj \leftarrow AT$

(4) $OP = JJ^* \leftarrow conj \leftarrow OP$

(5) $AT = NN^* \leftarrow comp \leftarrow AT$

Method

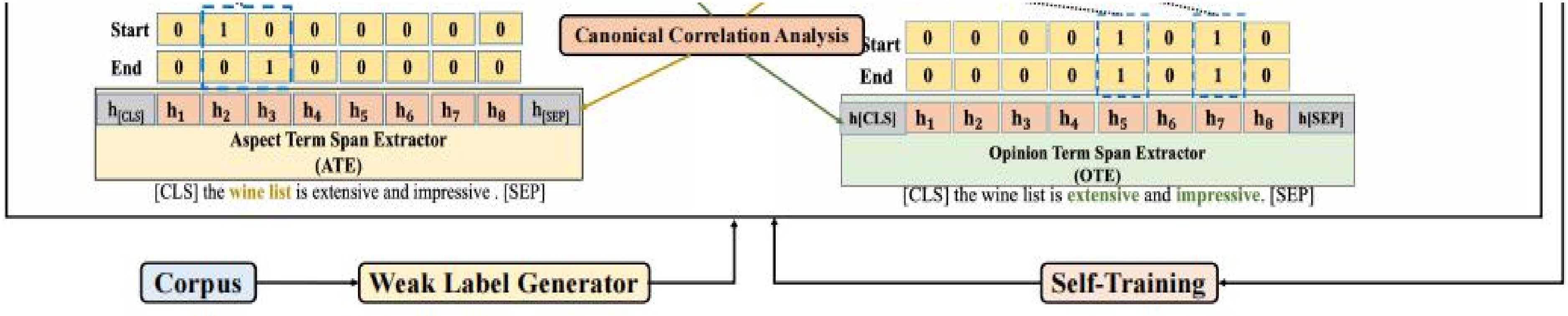


Figure 1: ODAO architecture

Aspect/Opinion Term Extractor

$$R = \{w_1, w_2, ..., w_k\}$$

$$H = \{h_{[CLS]}, h_1, h_2, ..., h_{[SEP]}\}$$

$$y_i = h_i * \mathcal{W}^T + b,$$

$$h_{i_s} = y_i[0],$$

$$h_{i_e} = y_i[1],$$

$$\hat{y}_i^s = \begin{cases} 1, & \text{if } h_{i_s} > 0; \\ 0, & else. \end{cases}$$

$$\hat{y}_i^e = \begin{cases} 1, & \text{if } h_{i_e} > 0; \\ 0, & else. \end{cases}$$

where $y_i \in \mathbb{R}^2$, $\mathcal{W} \in \mathbb{R}^{2*d_h}$, and $b \in \mathbb{R}^2$,

$\mathcal{W}$ and b are initialized randomly from $\mathcal{U}(-\sqrt{f}, \sqrt{f})$, where $f = \frac{1}{d_h}$

Experiments

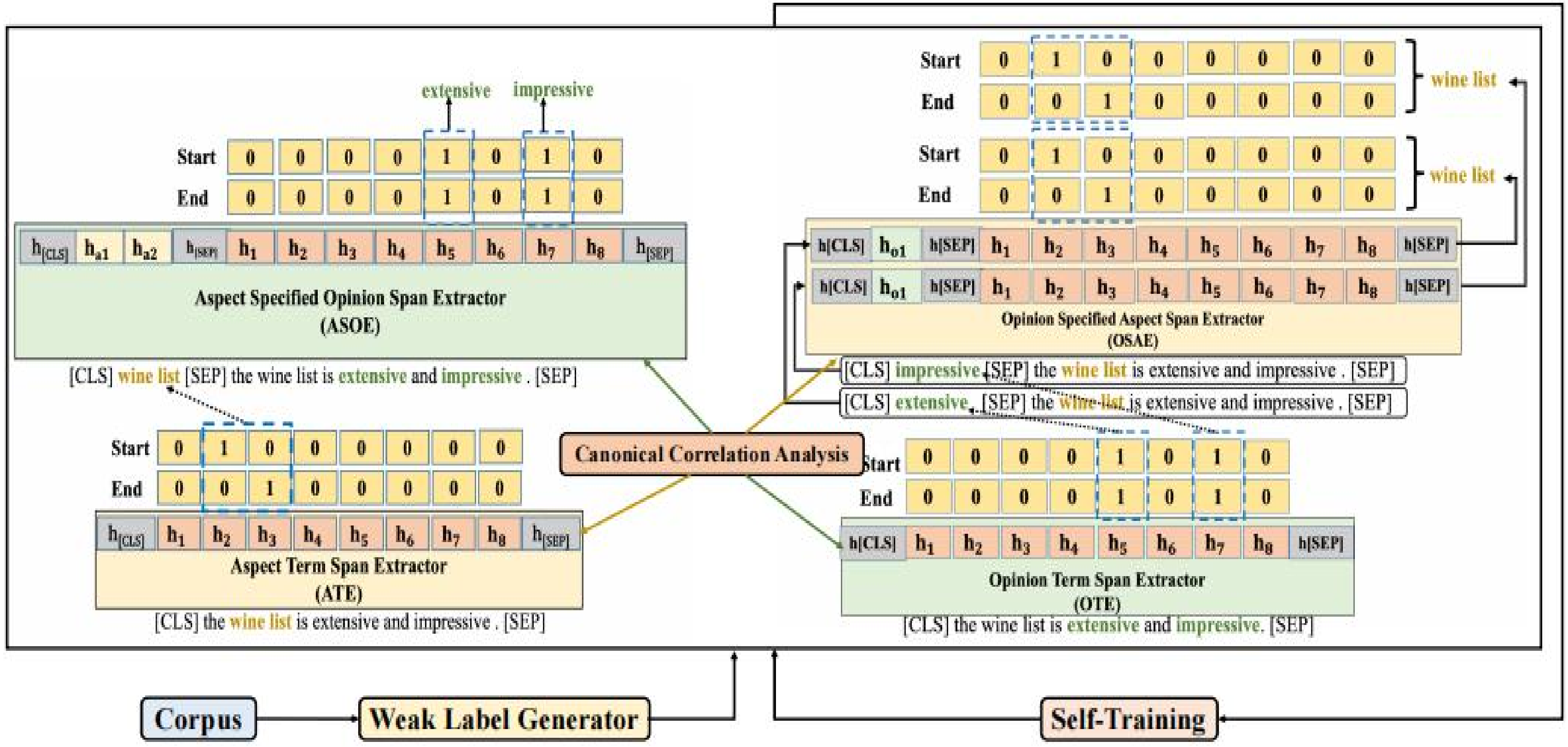


Figure 1: ODAO architecture

Aspect Opinion Pair Extractor

$$R = \{w_1, w_2, ..., w_k\}$$

$$A_p = \{AT_1, AT_2, ...\}$$

$$AT_i = \{a_1, .., a_q\}$$

$$I = \{[CLS], a_1, ..a_q, [SEP], w_1, w_2, ..., w_k, [SEP]\}$$

$$H = \{h_{[CLS]}, h_{a_1}, h_{a_2}, ..., h_{[SEP]}, h_1, ..., h_{[SEP]}\},$$

$\hat{y}_i^s$ $\hat{y}_i^e$

$$\mathcal{L}_{ASOE} = \frac{\mathcal{L}^s_{ASOE} + \mathcal{L}^e_{ASOE}}{2} = \frac{\sum_{i=1}^{N'} \sum_{sp \in \{s,e\}} BCE(\hat{y}_i^{sp}, y_i^{sp})}{2}$$

$$\mathcal{L} = \mathcal{L}_{ATE} + \mathcal{L}_{OTE} + \mathcal{L}_{ASOE} + \mathcal{L}_{OSAE}.$$

Method

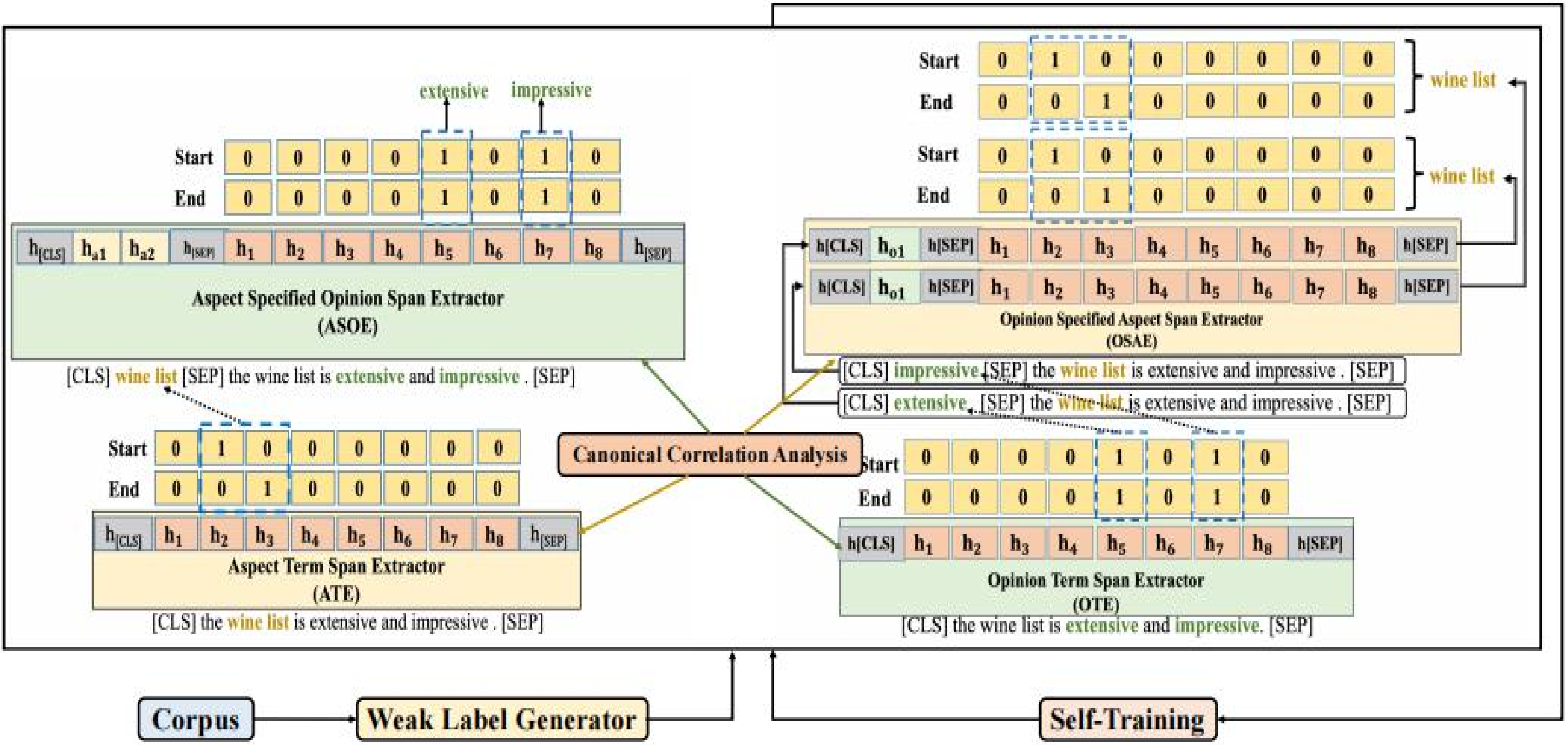


Figure 1: ODAO architecture

Early Stopping

The weakly labeled training data $\mathcal{D}'_{labeled}$ is biased and noisy. Motivated by [15], we employ early stopping to prevent the model from over-fitting to the label noise.

$$\rho_1 = corr(u^\top H_{ATE}, v^\top H_{OSAE}).$$

$$(u', v') = \underset{u,v}{\operatorname{argmax}}\ corr(u^\top H_{ATE}, v^\top H_{OSAE}),$$

$$\rho_1 = \frac{u^\top \Sigma_{AO} v}{\sqrt{(u^\top \Sigma_{AA} u)(v^\top \Sigma_{OO} v)}}.$$

$$\rho = \frac{\sum_M (\rho_1 + \rho_2)}{M},$$

where M refers to the number of reviews in the weakly labeled train set $\mathcal{D}'_{labeled}$.

Method

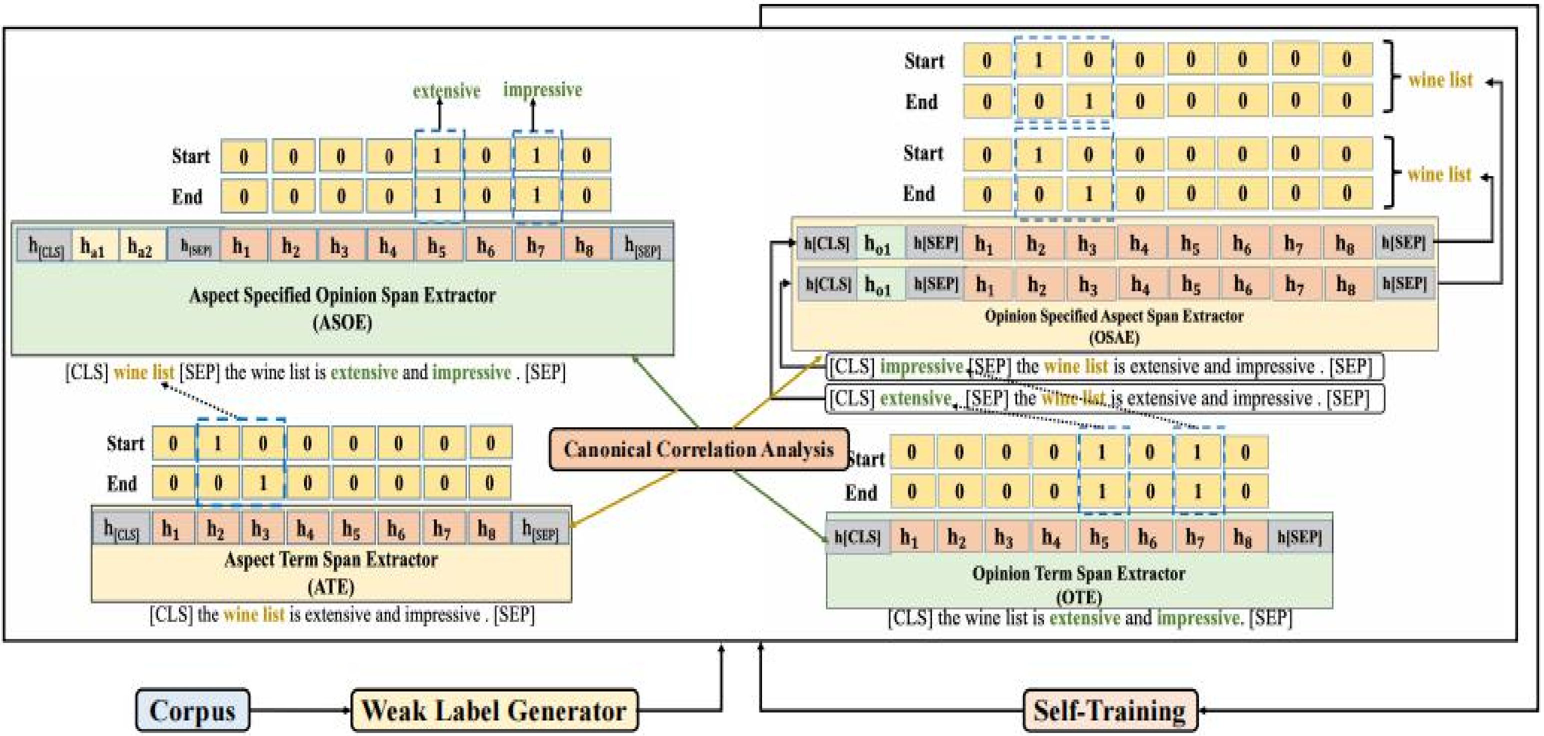


Figure 1: ODAO architecture

Self-Training

The weakly labeled train set $\mathcal{D}'_{labeled}$ constrains the proposed model performance due to the low coverage of the weak label generator rules. Furthermore, the bias in $\mathcal{D}'_{labeled}$ can also influence model training.

$$\gamma_R = A'_{ATE}\Delta A'_{OSAE} + O'_{OTE}\Delta O'_{ASOE}, \tag{9}$$

where $A\Delta B = (A - B)\bigcup(B - A)$ is the symmetric difference of two sets. All the modules agree on the predictions for a review R if $|\gamma_R| = 0$, and such reviews are considered to be correctly predicted.

Experiments

Table 2: Statistics of the Datasets

Datasets	$\mathbb{S}_{14}l$		$\mathbb{S}_{14}r$		$\mathbb{S}_{15}r$		$\mathbb{S}_{16}r$	
	Train	Test	Train	Test	Train	Test	Train	Test
#sentences	3045	800	3041	800	1315	685	2000	676
#aspects	2359	653	3693	1134	1205	542	1757	622
#opinions	2500	677	3512	1014	1217	516	1381	475

Table 3: Statistics of Fan et al. [7] datasets

Datasets	$\mathbb{S}_{14}l$		$\mathbb{S}_{14}r$		$\mathbb{S}_{15}r$		$\mathbb{S}_{16}r$	
	Train	Test	Train	Test	Train	Test	Train	Test
#sentences	1158	343	1627	500	754	325	1079	329
#pairs	1634	482	2643	865	1076	436	1512	457

Experiments

Table 4: Results of *ATE* task from *ATE* module on SemEval dataset. We report the span-level F_1 scores on the test sets. Results of the baselines are reported from their original papers. - refers to unpublished results as of the date of writing (Feb. 2022).

Method	Human Effort	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
RINANTE	Gold Annotation	80.16	86.45	69.90	-
QDSL		84.27	87.85	77.72	83.34
PSTD		**86.91**	88.75	75.82	82.56
DeepWMaxSat		81.33	85.33	-	73.67
FS-ODAO		85.93	**88.77**	**83.39**	**86.15**
ABAE	None	32.9	40.2		
LCC+GBC		36.1	41.2		
GMTCMLA	Sample Annotation	56.08	76.51	61.75	-
AutoNER	Dictionary	65.44	-	-	-
DP	Rule Design	19.19	38.72	27.32	-
ODAO		**76.14**	**80.73**	**80.72**	**79.24**

Table 5: Results of *OTE* task from *OTE* module on SemEval dataset. We report the span-level F_1 scores on the test sets. Results of the baselines are reported from their original papers. - refers to unpublished results as of the date of writing (Feb. 2022).

Method	Human Effort	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
RINANTE	Gold Annotation	81.96	85.67	72.09	-
DeepWMaxSat		80.34	85.73	-	79.67
DeepLogic		79.32	84.37	-	78.89
FS-ODAO		**85.47**	**87.23**	**84.56**	**88.43**
GMTCMLA	Sample Annotation	67.10	78.70	64.37	-
DP	Rule Design	55.29	65.94	46.31	-
ODAO		**77.82**	**79.57**	**82.56**	**81.26**

Experiments

Table 6: Results of *AOPE* task from *ASOE* and *OSAE* module on SemEval dataset. We report the span-level F_1 scores on the test sets. Results of the baselines are reported from their original papers. - refers to unpublished results as of the date of writing (Feb. 2022).

Method	**Human Effort**	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
QDSL	Gold Annotation	70.20	78.05	71.22	77.28
SDRN		67.13	76.48	70.94	-
SpanMlt		68.66	75.60	64.48	71.78
FS-ODAO		**90.04**	**89.89**	**87.18**	**90.06**
ODAO	Rule Design	81.75	83.02	83.93	81.41

Experiments

Table 7: Ablation Study results showing F_1 score for span-level *ATE* task from *ATE* module on SemEval dataset.

Methods	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
ODAO	76.14	80.73	80.72	79.24
-Pair Extraction Modules	50.13	57.53	60.86	60.71
-Self Training	62.06	72.19	72.13	71.0

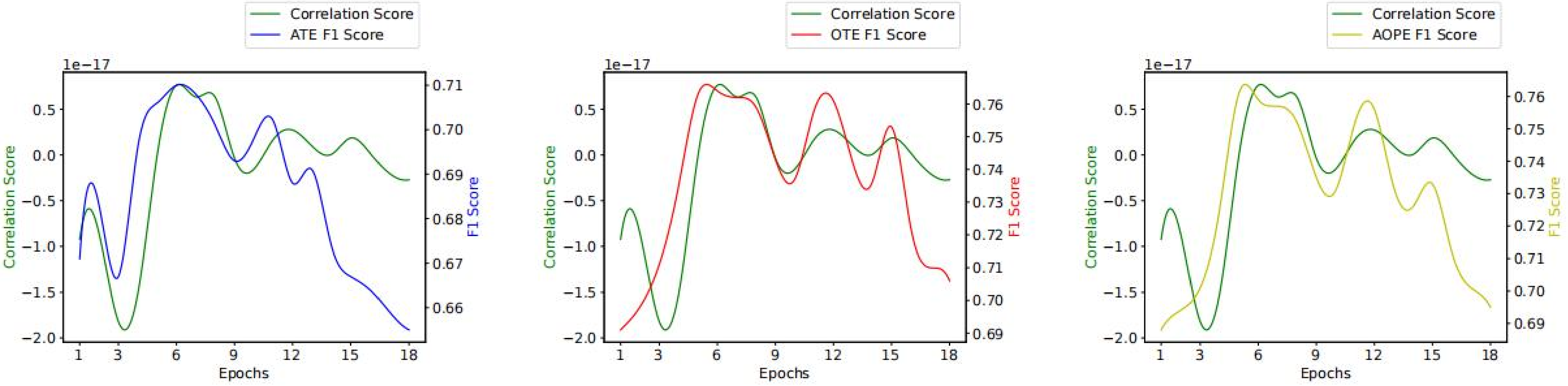
Table 8: Ablation Study results showing F_1 score for span-level *OTE* task from *OTE* module on SemEval dataset.

Methods	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
ODAO	77.82	79.57	82.56	81.26
-Pair Extraction Modules	72.75	75.63	78.45	77.56
-Self Training	73.30	76.70	77.37	76.3

Table 9: Ablation Study results showing F_1 score for span-level *AOPE* task combining *ASOE* and *OSAE* module on SemEval dataset.

Methods	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
ODAO	81.75	83.02	83.93	81.41
-Self Training	70.64	76.21	76.65	76.09

Experiments



(a) CCA score and ATE F1 score (b) CCA score and OTE F1 score (c) CCA score and AOPE F1 score

Figure 3: CCA score and model performance on different tasks

Experiments

Table 10: Case study of reviews with complex aspect-opinion relation

Review	Model Predictions	Ground Truth
i recommend the black roasted codfish , it was the best dish of the evening .	**ATE**: [black roasted codfish, dish], **OTE**: [recommend, best], **AOPE**: [(black roasted codfish, recommend), (dish, best)]	**ATE**: [black roasted codfish, dish], **OTE**: [recommend, best], **AOPE**: [(black roasted codfish, recommend), (dish, best)]
– i ca n't say enough about this place .	**ATE**: [place], **OTE**: [null], **AOPE**: [(null, null)]	**ATE**: [place], **OTE**: [null], **AOPE**: [(null, null)]
it 's fast , light , and simple to use .	**ATE**: [use], **OTE**: [fast, light, simple], **AOPE**: [(use, fast), (use, light), (use, simple)]	**ATE**: [use], **OTE**: [fast, light, simple], **AOPE**: [(use, fast), (use, light), (use, simple)]
i can highly recommend their various saag and paneer and korma .	**ATE**: [saag, paneer, korma], **OTE**: [recommend], **AOPE**: [(saag, recommend), (paneer, recommend), (korma, recommend)]	**ATE**: [saag, paneer, korma], **OTE**: [recommend], **AOPE**: [(saag, recommend), (paneer, recommend), (korma, recommend)]

Thanks